


Cloud computing based unsupervised fault diagnosis system in the context of Industry 4.0

Sistema de diagnóstico de falhas não supervisionadas baseado em computação em nuvem no contexto da Indústria 4.0

Amr Mohamed Ali¹, El-Adl Mohamed¹, Soumaya Yacout², Yasser Shaban¹ 

¹Helwan University, Department of Mechanical Design Engineering, Faculty of Engineering, Cairo, Egypt.
E-mail: yasser.shaban@polymtl.ca

²École Polytechnique, Department of Mathematic and Industrial Engineering, Montréal, Québec, Canada

How to cite: Ali A. M., Mohamed E-A, Yacout S., Shaban Y. (2020). Cloud computing based unsupervised fault diagnosis system in the context of Industry 4.0. *Gestão & Produção*, 27(3), e5378. <https://doi.org/10.1590/0104-530X5378-20>

Abstract: New online fault monitoring and alarm systems, with the aid of Cyber-Physical Systems (CPS) and Cloud Technology (CT), are examined in this article within the context of Industry 4.0. The data collected from machines is used to implement maintenance strategies based on the diagnosis and prognosis of the machines' performance. As such, the purpose of this paper is to propose a Cloud Computing Platform containing three layers of technologies forming a Cyber-Physical System which receives unlabelled data to generate an interpreted online decision for the local team, as well as collecting historical data to improve the analyzer. The proposed troubleshooter is tested using unlabelled experimental data sets of rolling element bearing. Finally, the current and future Fault Diagnosis Systems and Cloud Technologies applications in the maintenance field are discussed.

Keywords: Remote Fault Diagnosis System (RFDS); Logical Analysis of Data (LAD); Cyber-Physical System (CPS); Pattern recognition; Industry 4.0; Cloud computing.

Resumo: Novos sistemas de monitoramento e alarme de falhas on-line, com a ajuda dos sistemas ciber-físicos (CPS) e da tecnologia em nuvem (CT), são examinados neste artigo no contexto da indústria 4.0. Os dados coletados das máquinas são utilizados para implementar estratégias de manutenção com base no diagnóstico e prognóstico do desempenho das máquinas. Assim, o objetivo deste artigo é propor uma Plataforma de Computação em Nuvem contendo três camadas de tecnologias que formam um Sistema Ciber-Físico que recebe dados não rotulados para gerar uma decisão on-line interpretada para a equipe local, além de coletar dados históricos para melhorar a analisador. O solucionador de problemas proposto é testado usando conjuntos de dados experimentais não identificados de rolamentos de corpos rolantes. Por fim, são discutidos os atuais e futuros sistemas de diagnóstico de falhas e a aplicação de tecnologias de nuvem no campo de manutenção.

Palavras-chave: Sistema de diagnóstico remoto de falhas (RFDS); Análise lógica de dados (LAD); Sistema físico cibernético (CPS); Reconhecimento de padrões; Indústria 4.0; Computação em nuvem.

1 Introduction

Without a doubt, maintenance is an important chain of the product lifecycle, since it costs as much as 60% to 70% of the total cost of the product lifecycle (Dhillon, 2006; Venkataraman, 2010). Although the maintenance cost is considerably high, the existing maintenance solutions are used in isolation without taking into consideration the real status of the machine components (Efthymiou et al., 2012). Due to the interlinked production systems, failures of the machine components lead to bottlenecks in the subsequent value-added processes of the manufactures and their customers (Denkena et al., 2012; Wang, 2013). Advanced manufacturers have introduced the use of remote diagnosis techniques to gain awareness of equipment status and identify failures beforehand (Mori et al., 2008). Remote Fault Diagnosis System (RFDS) is defined as a troubleshooter used to diagnose a subject (system, subsystem or component) at distance (Bidgoli, 2010). RFDS has three major layers: I) sensors installed on the monitored subject used to collect the data from the diagnosed system; II) a communication network that conveys the collected data from the diagnosed system to; III) the remote diagnosis center where the data of the information system process occurs and a decision is made based on the generated information of the subject status. The diagnosis center includes real decision making management software that uses the historical data of the system performance to diagnose, isolate, identify the root cause and monitor the performance of the system. RFDS is commonly used in medical equipment, aircraft, and military equipment. Figure 1 illustrates the three layers of RFDS and explains the communication between them (Sanchez et al., 2011).

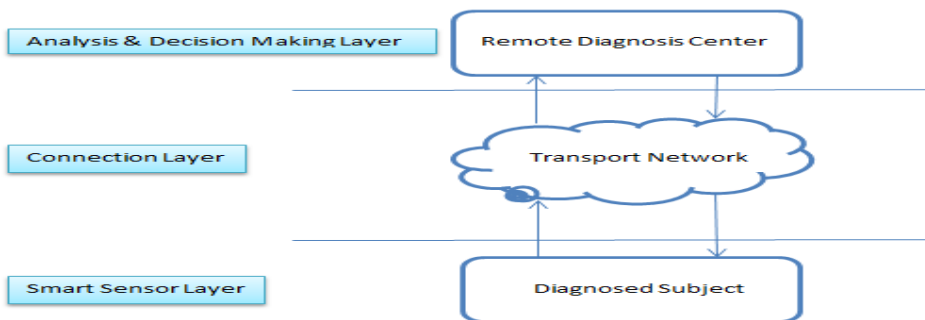


Figure 1. RFDS Layers.

The default setup of the remote diagnosis system exchanges the data with a service hotline, where human experts give their technical advice via communication methods, such as emails and video calls (Wang et al., 2004; Sekar et al., 2011). The advanced setup of the remote diagnosis system, on the other hand, is the automated failure diagnosis system, which uses an offline analysis technique classifying the behavioural change based on the historical data of the subject; then it sends the necessary action to the local maintenance team and the spare parts support (Mourtzis et al., 2017; Zhao et al., 2010).

2 Literature review

Many of the most popular diagnosis and decision making support have already been applied for the automated remote diagnosis system in rotating machinery. For example, Mourtzis et al. (2017). created an advanced monitoring system by using a cloud-based platform supported by argument reality to monitor the machine tools, calculate remaining useful time and perform accurate and quick maintenance action. Zhao et al. (2010) described an expert enterprise monitoring and remote fault diagnosis system based on web services, visual studio and extensible mark-up language technologies. This proposed architecture enhances elasticity and scalability while decreasing the time needed to develop the maintenance system. Jantunen et al. (2018) described the concept of advanced condition-based maintenance with the aid of cloud computing. This study aims to reinforce the idea that a maintenance system supported by CT and CPS can increase the overall equipment efficiency and reduce maintenance costs. Campos et al. (2016) highlighted data management in advanced maintenance systems, especially when emergent technologies such as CT and big data are used. The use of big data and CT has been discussed and the element of cloud-based big data system features has been highlighted. Monostori et al. (2016) studied the development of computer science and information technologies used to build a cyber-physical system and its implementation in the manufacturing process. The root of the cyber-physical production system from manufacturing technology and the main research challenges are highlighted.

Additionally, Kirubashankar & Krishnamurthy (2013) implemented a novel web-based expert maintenance support system for the Distributed Control System (DCS) on chemical plants, which are one of the vital components of any manufacture. The collected information and critical alarms are stored and compared with the reference functional database instantly; then they are sent to the maintenance engineer for online diagnosis. Sanz-Bobi et al. (2012) developed a remote diagnosis system capable of performing continuous and automatic monitoring of the basic parameters of photovoltaic solar power plants to detect anomalies. A prototype of the proposed system and architecture is introduced for the photovoltaic solar power plant. On the other hand, the concept of cyber-physical systems was proposed in Lee et al. (2015). This concept is based on five levels: I) connection level; II) data to information conversion level; III) cyber level; IV) cognition level; and V) configuration level. This new trend of the cyber-physical system is transforming the manufacturing industry to the next generation industry of 4.0. The main advantage of CPS is the interface between machines, which is similar to social networks. Holgado et al. (2016) studied the developing trends in the remote maintenance, such as diagnosis and prognosis tools, cloud-based tools, simulation tools, location, and tracking tools, aiming to explain the value of using the new trend in service provision. While Wang et al. (2017) presented a new paradigm of cloud predictive maintenance based on mobile agent technology. The mobile-agent based paradigm reduces data transmission and enhances system flexibility. Also, a proactive collaborative maintenance CPS project is proposed to building a platform for proactive maintenance of industrial machines Jantunen et al. (2016). Mourtzis et al. (2016) presented a preventive maintenance approach integrated into a monitoring framework. The developed approach was built by using CT. The main advantage of the proposed system is the distribution of real-time information related to the machine tools to the operators.

Based on the former literature survey, this paper's originality arises from its aim to build a cloud computing platform for diagnosis of the machine's performance; and the exploration of unlabelled data to generate interpretable patterns to be used in

maintenance decision making. The combining of machine learning and communication technology brings us to a new age of industrial engineering.

In Section 3 below, the proposed methodology is introduced, while in Section 4 the remote diagnosis unlabelled system architecture is proposed; whereas in Section 5, a case study is presented. Finally, Section 6 exhibits the discussion and conclusion of this paper.

3 Methodology

Several types of data collection methods can be used to detect the status of the monitored subject. The signals can be vibration signals, thermal imaging, oil particle signals... etc. The most common technique of fault detection for rotating machines is the vibration signal. In the following section, the vibration signal analysis is illustrated. This section presents an unsupervised machine learning technique component of three layers to discover and understand the hidden correlation between the vibration signals. The first layer is vibration signal features analysis, the second layer is clustering and the third layer is a pattern-based machine learning technique called Logical Analysis of Data (LAD).

3.1 Vibration signal features analysis

The first analytical step of the vibration signal is feature extraction. Those forms can be used to diagnose the failure status in the rotating system. However, in the three forms, there are a lot of hidden data that can be extracted to improve the accuracy of the diagnosis technique. The advantages of the hidden data extraction techniques and the disadvantages are studied in Lakis (2007), whereas the following table illustrates the statistical features which are commonly used in many research as stated in Mortada et al. (2011). The time domains feature calculated from the vibration signal are descriptive statistics are shown in Table 1.

Table 1. Vibration signal feature formula.

Feature Name	Formula
Range (R)	$Range = Max(y_i) - Min(y_i)$
Mean Value (M)	$Mean = \frac{\sum_{i=1}^N y_i}{N}$
Standard Divination (D)	$STDE = \sqrt{\frac{\sum (y_i - \bar{y})^2}{(N - 1)}}$
Skewness(S)	$Sk = \frac{n}{(n - 1)(n - 2)} \sum \left(\frac{y_i - \bar{y}}{STDE} \right)^3$
Kurtosis (K)	$Ku = \frac{n(n + 1)}{(n - 1)(n - 2)(n - 3)} \sum \left(\frac{y_i - \bar{y}}{STDE} \right)^4 - \frac{3(n - 1)^2}{(n - 2)(n - 3)}$
Root Mean Square (RM)	$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}$
Crest Factor (C)	$CF = \frac{Max(y_i)}{RMS}$

"N" is the length of data, while "y" is the amplitude of vibration signal.

The other challenge faced is that there is no scientific node between healthy and unhealthy classes. The new improvement on LAD is using a clustering technique to convert the unlabeled data into a labeled one to help LAD discover the hidden patterns inside the data. In the following section, we introduce the clustering algorithm.

3.2 The clustering analysis

After analyzing the vibration signals we use the extracted feature of this signal to find the node between the classes to illustrate the information inside the vibration signal. The Clustering Analysis aims to characterize homogenous objects into a set of classes dealing with very large-scale and complex data sets. We use K-mean clustering algorithm to find the best clusters. K-mean is the most common algorithm used in signal processing. The clustering process is done by minimizing the sum of squared distances between data and the corresponding cluster centroid. This technique works through four steps. The first step in how the algorithm work is random creating of K centroids based on the initial defined of K. the second step is to allocate each point of the dataset to the nearest centroid. The third one has recomputed the mean of every data point to the centroid's cluster to reduce the intracluster.

The final step is to reiterate the previous steps (2, 3) to find the best criteria. The following Equation 1 explains the K-mean algorithm formula (Zhang, 2008; Kanungo et al., 2002).

$$J(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} (\|x_i - v_i\|)^2 \quad (1)$$

$(\|x_i - v_i\|)$ is the Euclidean distance between x_i and v_i . c_i is the number of data points in the first cluster, whereas c is the number of cluster centers.

The main advantage of the K-mean is that it is fast, robust, easier to understand and gives the best result when data sets are distinct or well separated from each other. In the following section, we use the proposed methodology to build the cyber-physical maintenance system.

3.3 Logical Analysis of Data (LAD)

LAD is a data-driven technique that allows for the discovery of the hidden patterns of phenomena based on pattern recognition. It is applied through two stages: training and testing stages. In the first stage, part of the data is used as training data to extract the strongest hidden patterns explaining the phenomena. To build the most accurate diagnosis model, the training data shall be well balanced to cover the healthy and failure classes, while the rest of the data shall be used to test the generated mode accuracy from the first stage. LAD was introduced at the Rutgers University Center for Operations Research (RUTCOR) in Hammer (1986). The comparison between LAD and the other techniques of machine learning are proposed in Yacout (2010). The main steps of the proposed technique are the binarization of data, pattern generation, and theory formation. The first step is the process of transforming data into a Boolean database. In our work, we use the binarization technique proposed in Boros et al. (2000). The proposed binarization technique is briefly explained in Shaban et al.

(2017). The pattern generation is the second step in the proposed technique; in this step, the key building block in LAD extraction is generated. In our work, MILP is used to generate the hidden patterns of the failure state of the rolling element bearing. The generated patterns cover at least one point of the observations and the strong pattern covers most observations.

The last step in the proposed technique is theory formation. In these steps, the LAD decision model is generated. It can be used in real-time remote maintenance systems and decision-making support. The equation explained in Mortada et al. (2011) generate a score ranging between -1 and 1. When the output of the presented function is a negative value, this means that the tested observation belongs to the negative class. On the other hand, if the result equals zero, this means that LAD cannot decide the class of the tested observation. ; In the follow we minutely discuss each of those steps.

3.4 Data binarization

In this step, the raw data transformed into binary data using a binarization technique that translates each numerical feature to a set of binary attributes. The binarization technique use ranks the distinct values that a numerical feature (u) takes in a data set in ascending order. The second step is deciding the cut point is inserted between every two values which belong to different classes. The average of the two values is calculated to find the cut-point, value. For each cut-point, a binary attribute b is formed as follow in Equation 2.

$$b(u) = \begin{cases} 1 & \text{if } u \geq \alpha \\ 0 & \text{if } u < \alpha \end{cases} \quad (2)$$

The number of cut points generated for each numerical feature is the same generated number of binary attributes that make up the binarization training set at the end of the binarization process. In the following section, the pattern generation step is explained.

3.5 Pattern generation

This step aims to discover the hidden patterns that divide between classed. In our study, the classes are positive class (defect) and negative class (healthy). In this paper, the pattern generation is proposed based on the formulation and solution of a mixed-integer and linear programming (MILP) problem (Linderoth & Lodi, 2010). The advantages of this approach than the described pattern generation methodology proposed in (Ryoo & Jang, 2009; Moreira, 2000) are the ability to control the discriminating power between the different classed through a user-defined parameter called the discriminating which explain the minimum number of patterns that must separate each observation to class c_i from those belonging to class c_j (Mortada et al., 2014). The second advantage of the proposed approach is no limit on the number of patterns generated by the pattern generation algorithm, which creates more patterns finding possibilities.

In this step the class patterns that create the decision model in the last step of the proposed approach. It started by creating an empty set of patterns $P_{i,j}$ for each pair of class (c_i, c_j) , where $i, j \in \{1, 2\}$. The pattern sets a relationship with the class pairs to

which they belong is governed by the matrix $R_l(K \times K)$ whose elements $r_{i,j}$ are calculated by the following Equation 3:

$$r_{i,j} = \frac{disc_l(c_i, c_j)}{m_i} \tag{3}$$

The factor of discriminating $l \geq 1$ determines the minimum number of patterns inset P_{ij} which separates each observation belonging to class c_i from those in class c_j . The generated pattern P is tested for all class pairs (c_a, c_b) where $a, b \neq i, j$, to decide on whether P can be placed in their corresponding pattern set P_{ab} . The pattern rules to be admitted to a set P_{ab} are controlled by the two user-defined values: the minimum positive coverage rate d_{min+} and the maximum negative coverage rate d_{max-} . The generated pattern is added to each set of patterns that satisfies the following coverage rate thresholds as follow in Equation 4 and 5:

$$d_a(p) = \frac{cov(p, c_a)}{m_a} \geq d_{min+} \tag{4}$$

$$d_b(p) = \frac{cov(p, c_b)}{m_b} \geq d_{max-} \tag{5}$$

The $cov(p, c)$ is the number of an observation belonging to a class c that covered by pattern p , m is the total number of observations (m) that belong to class c . The resulting pattern has the highest positive coverage rate and the least possible degree to increase its generalization power.

The Boolean pattern vector $w(w_1, w_2, w_3, \dots, w_{2q})$ has a double number of binary attributes q that generate in the binarization step. Naturally, a pattern cannot include both the literal and negation x_j, \bar{x}_j at the same time. The constraint is illustrated as follow in Equation 6:

$$w_j + w_{q+j} \leq 1 \quad j = 1, 2, \dots, q \tag{6}$$

The MILP solution is a pattern p of degree δ . The pattern composition can be summarizing from vector and the coverage form vector y . The following Equation 7 illustrates the optimal solution:

$$\min_{w, y, \delta} \sum_{i \in S^+} y_i + g\delta \tag{7}$$

The three major constrains sets of MILP problem to generate the optimized pattern of the objective function. The first constraint that the resulting pattern must be able to cover a positive observation $i \in S^+$ and not required to cover all the observations in c . This constrains illustrate as follow in Equation 8:

$$\sum_{j=1}^{2q} a_{ij}w_j + qy_i \geq \delta \forall_i \in S^+ \tag{8}$$

The second constraint in the optimization problem is that the positive pattern should be not covering any negative observations. This constraint is shown as follow in Equation 9:

$$\sum_{j=1}^{2q} a_{ij}w_j \leq \delta - 1 \forall_i \in S^- \tag{9}$$

The last constraint in the MILP problem is it must not generate a pattern that has been generated in previous runs. The mathematical formulation which achieves this constrain as follow in Equation 10:

$$\sum_{j=1}^{2q} v_{k,j}w_j \leq \delta_k - 1 \forall_i \in V \tag{10}$$

where V is the set containing the Boolean pattern vectors of all the generated patterns in the set P_{ij}. The MILP problem is illustrated below, where m is the number of observation in S⁺ as follow in Equation 11 and 12:

$$\min \sum_{w,y,\delta_i \in S^+} y_i \tag{11}$$

$$s.t. \begin{cases} \sum_{j=1}^{2q} w_j = \delta \\ 1 \leq \delta \leq q \\ w \in \{0,1\}^{2q} \\ y \in \{0,1\}^m \end{cases} \tag{12}$$

3.6 Theory formation

Theory formation is the final step of the LAD decision-making model where the generated patterns in the previous step are used to form a decision function called the discriminant, such as the one given in Equation 13. This equation used to compute a score ranging between -1 and 1. In case the output of the function has a positive value, this indicates that this signal is for a defected element. If the equation output is negative this means that the element is in healthy stat. Zero output means that the signal not enough to detect to which class this signal belongs to each class (Mortada et al., 2014).

$$\Delta(O) = \sum_{i=1}^{N^+} \alpha_i^+ P_i^+(O) - \sum_{i=1}^{N^-} \alpha_i^- P_i^-(O) \tag{13}$$

where N is the number of generated patterns, $P_{(O)}$ is equal to 1 if pattern covers observation O , otherwise is equal to zero and α is the weight of pattern.

4 Architecture of unlabelled remote diagnosis system

For the industrial process and maintenance of machines, CT and data analysis are important technologies and are gradually becoming the trend in maintenance innovation. The major advantages of CT are cost-effectiveness, unlimited storage space, data recovery, accessibility, scalability, and security. To achieve near-zero break down, machines need to be continuously monitored during their life cycle. This needs scalable storage space and analysis. A detailed architecture of maintenance for the cyber-physical system that consists of three layers, namely, sensor layer, connection layer, and analysis and decision making layer is introduced.

- 1) **Sensor Layer:** this layer consists of sensors providing multisource unlabelled data from machines. The sensor has the capability, such as wireless communication, enabling the data transfer through the internet to the analysis and decision making layer. Sensors already support dynamic, real-time, optimized and self-organized industry processes. Intelligent sensors report the real statuses of the machine and machining process, turn these into digital data, and share them automatically with the process controller;
- 2) **Connection Layer:** this layer consists of distributed file storage and parallel computing. By using CT, the connection between the physical side and the cyber side occurs;
- 3) **Analysis and Decision-Making Layer:** this layer includes the analyzer software that transfers the unlabelled data into useful information and decisions, which are sent back to the machine and the local maintenance team. This software was built by using historical data analysis, which generates governor patterns.

The objective of this paper is to build Unlabelled Cyber-Physical Maintenance System Software (maintenance-CPS) by using an interpretable methodology to generate automatically and without any human interference the status of the machine components (whether healthy or unhealthy); identify the corrective action needed; and, alarm the local maintenance team to interact. Moreover, the proposed platform is built to be the connection point between all maintenance management elements consisting of: monitored machines, local maintenance teams, and spare parts supplier, to overcome the problem of work at a distance; and CT is used to connect its layer. To this end, it is very important to identify Cloud services. Cloud services are internet-based services which can offer all the things that a computer system provides. Due to the continuous increase of data collation, there is an overwhelming flow of data (Villars et al., 2011). Cloud technologies have many advantages of cloud computing including virtualized resources, parallel processing, security and data service integration with scalable data storage. It does not only minimize the cost and restrictions but also provides reduced infrastructure maintenance cost; efficient management and user access control (Hashem et al., 2015). Additionally, Cloud services provided by vendors, such as Dropbox, IBM, Microsoft Azure, Google Cloud and Amazon AWS, support 24/7 operations (Avram, 2014). CT solves the problems faced when using traditional computing. Simply a required capacity of computer service needs can be rented from one of the service providers mentioned above. The virtualized machine can provide the best performance with low cost without the

need for a high capital cost for servers, power, cooling, software and hardware upgrades. In this present study, the researchers use Dropbox Cloud storage service as a distributed file storage. The storage capacity of the prototype used for this research is (2 Gigabyte). The benefits of the proposed platform are: adjustable storage capacity; easy and secure access; as well as, fast and efficient synchronize technology.

The proposed system starts with an offline unlabelled data analysis to generate governor patterns. The generated patterns are transmitted to an online decision making. The architecture of the unlabelled cyber-physical maintenance system is divided into three layers: the first layer is the sensor layer, which uses an intelligent sensor that connects to the internet to send the collected data to the analysis and decision-making layer passing through the connection layer. The second layer is the connection layer using CT to save and transfer the data between the different system layers via the internet. The third layer is the analysis and decision layer that is divided into two steps: the first step is unlabelled to labelled data transformation by using the K-mean algorithm. This step enables LAD to discover the hidden patterns in the vibration signal. The historical data is used to help offline LAD to generate constraints, which are used to build the cyber-physical system to continuous monitor, diagnosis, and control. The second step is building the cyber-physical center, which decides to be sent back by the Cloud to the machine control unit, then being transmitted to both the maintenance team to take the necessary action accordingly and the spare parts supplier to prepare the spare parts needed to help the maintenance team to finish the maintenance task in the shortest time possible, reducing the maintenance cost and increasing the production rate. Furthermore, the generated cyber-physical system sends back a corrective action to the machine to overcome the fault or to shut down the machine to prevent catastrophic failure. Figure 2 illustrates the proposed architecture of the unlabelled cyber-physical system.

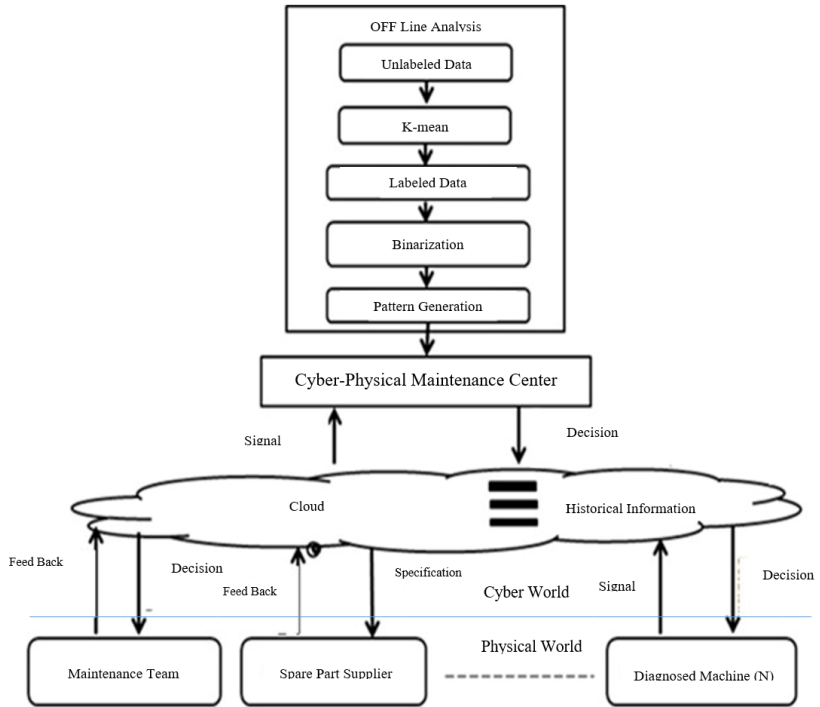


Figure 2. Architecture of Unlabelled Cyber-Physical Maintenance System.

The starting point of the proposed system is the sensors, which take a signal of the monitored machine every certain time; adjust it according to the application including detection or measurement parameters enable them to be dynamically adapted to tasks as necessary; and finally, send it to the Cloud. The collected unlabelled data go to the online remote diagnosis center to convert it to useful information and determine the status of the machine. A decision is taken, either the machine is in a healthy state, then the system records their status in a report; or the machine is in an unhealthy state, upon which the system sends an alarm and corrective action to the local team and spare parts supplier. The corrective action to replace the defected part is sent to the local team. Also, the specification of the needed part is sent to the spare part supplier to prepare the required element. The historical data stored in the Cloud is used to improve the accuracy of the system. Figure 3 illustrates the decision making procedure schematic of unlabelled cyber-physical maintenance system (Maintenance-CPS).

In Figure 3 the solid line is an indication for wired connection while the dashed line is connected via the internet. Besides, corrective actions are sent to the machine actuator to control the operation factor.

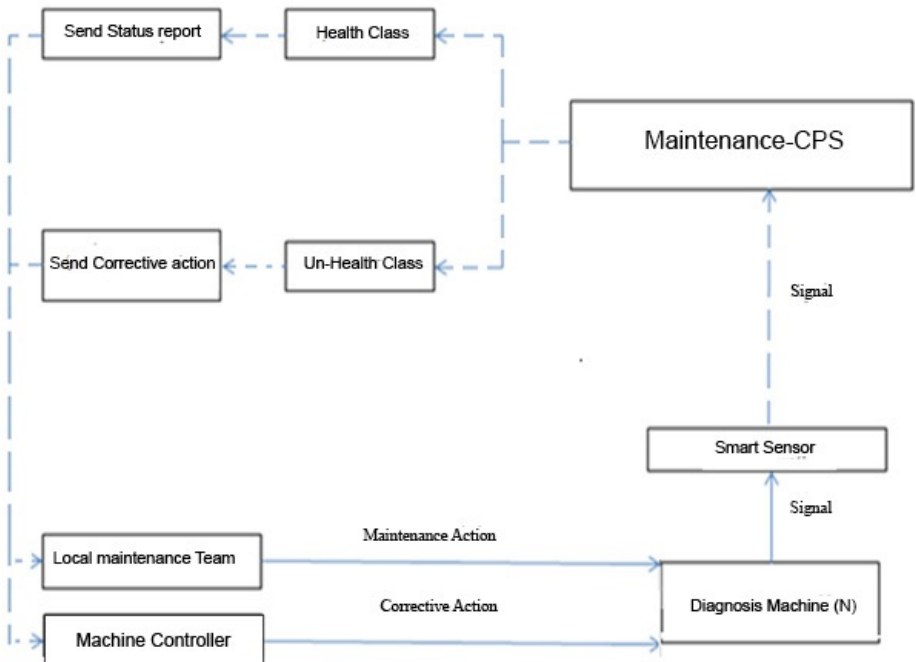


Figure 3. Decision Making Procedure Schematic.

5 Case study

To test the proposed system, a simulation model of the unlabelled cyber-physical maintenance system of rolling element bearing failure is developed. The tested data is extracted from a run to failure data sets of signals obtained from the National Science Foundation’s Industry/University Cooperative Research Centre for Intelligent Maintenance Systems (IMS) through the NASA prognostic data repository shown in Figure 4 below. The test rig setup consists of a constant speed motor (2000RPM) coupled to a shaft supported with four rolling element bearing. The four types of bearings are Rexnord ZA-2115 double-row bearings with 16 rollers in each row, 71.5 mm pitch diameter, 8.407 mm roller bearings and a tapered contact angle of 15.170 (Lee et al., 2015). A high sensitivity Quartz ICP accelerometers PCB 353B33 was installed on the bearing housing. The card was used to collect a snapshot of the vibration signal every ten minutes with 20 KHz sampled signals until a failure occurred. This test was carried out for seven days. Table 2 illustrates a sample of feature extraction using MATLAB Software.

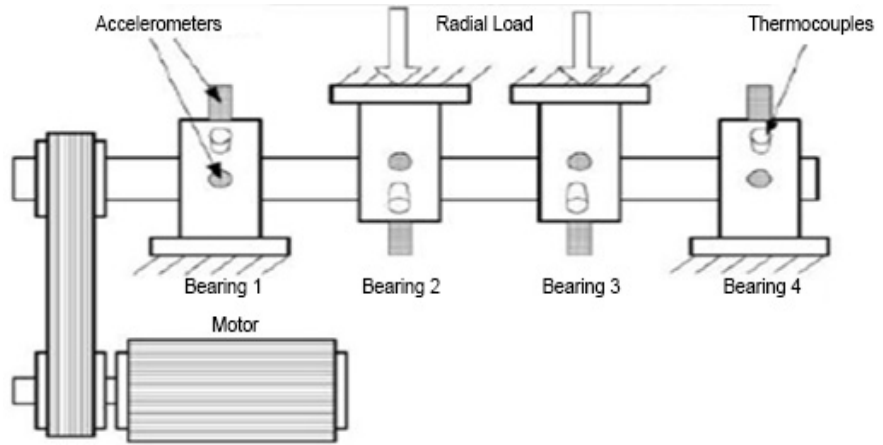


Figure 4. Bearing test rig and sensor placement illustration (Qiu et al., 2006).

Table 2. A sample of the unlabelled extracted feature.

Date	Time	Range	Mean	STDE	SK	KU	RMS	CF
12/2/2004	10.32	0.840	-0.010	0.073	0.084	0.629	0.074	6.120
12/2/2004	11.02	1.184	-0.002	0.079	0.041	1.159	0.079	7.723
12/2/2004	12.02	0.722	-0.002	0.076	0.046	0.326	0.076	4.992
12/2/2004	13.02	0.725	-0.002	0.079	-0.013	0.553	0.079	4.500
12/2/2004	14.02	0.742	-0.002	0.079	0.038	0.385	0.079	4.801
19/2/2004	02.02	1.489	-0.003	0.207	-0.161	0.169	0.207	3.286
19/2/2004	03.02	1.995	-0.002	0.235	0.090	0.447	0.235	4.144
19/2/2004	04.02	3.940	-0.001	0.326	-0.122	1.597	0.326	5.096
19/2/2004	05.02	9.998	0.003	0.672	-0.766	14.114	0.672	7.436
19/2/2004	06.02	7.197	-0.002	0.484	-0.377	4.893	0.484	7.236

Where STDE is standard deviation; SK is Skewness; KU is Kurtosis; RMS is Root mean square; and CF is crest factor. The proposed clustering technique was used to find the nodal point that identifies the status of the bearing. Then the K-mean clustering step was implemented by using the Weka data mining software (Hall et al., 2009). In the following Figure 5 and Table 3, the clustering results are illustrated by using K-mean algorithm.

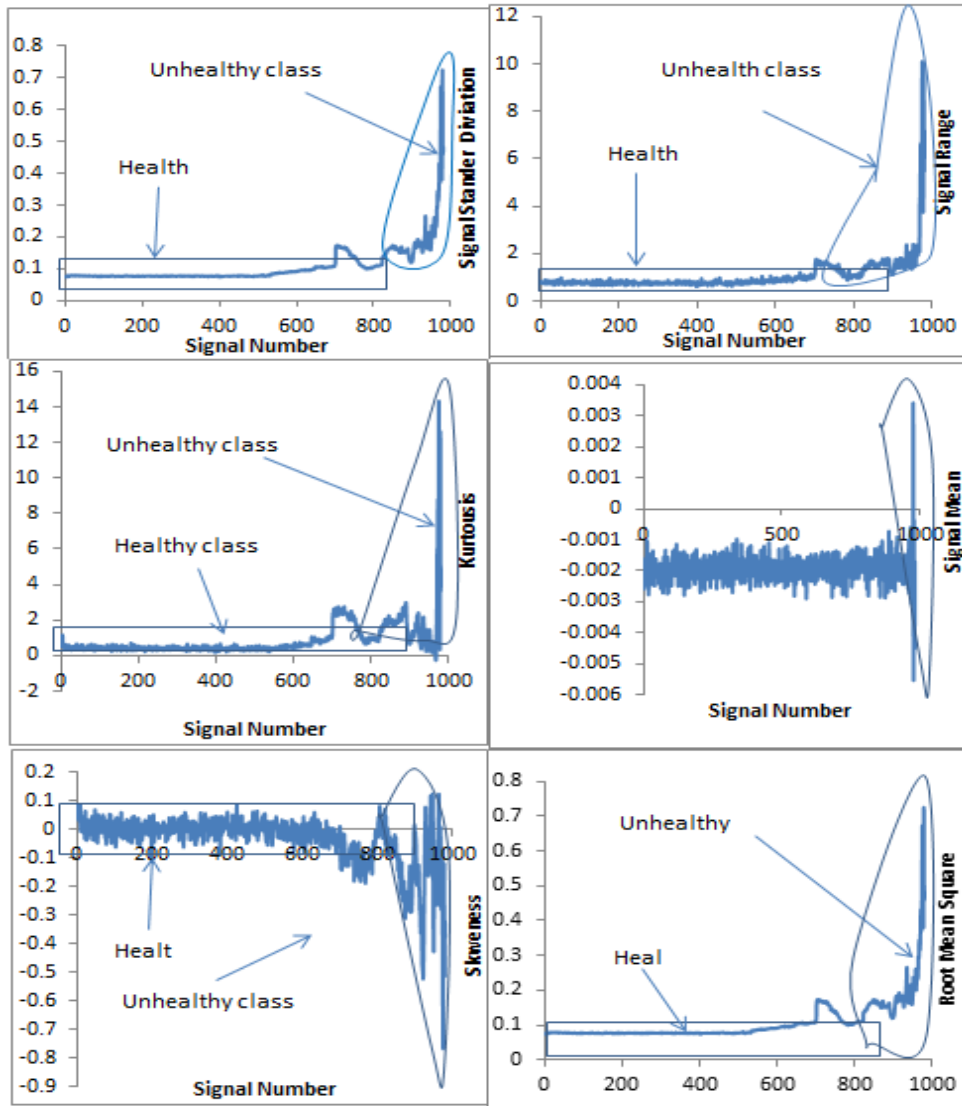


Figure 5. K-mean node presentation using Signal statistical analysis.

Table 3. K- mean clustering results.

No.	Class Name	Quantity
1	Healthy class	751
2	Unhealthy class	206

As exhibited in Table 3 above, the new data after clustering was used to detect the characteristic patterns for machine operation status that lead to a healthy or unhealthy condition. LAD generates positive patterns that discover the unhealthy condition and negative patterns in the condition of health status. The software cbmLAD (Yacout et al., 2012) is trained by using the data obtained from the previous step. Table 4 shows a sample of the positive characterization patterns and the negative characterization patterns found by cbmLAD software.

Table 4. Sample of generated positive and negative patterns.

N.	Positive Patterns	Negative Patterns
1	$R > 1.169, K > 0.1072, RM > 4.917$	$D > -0.053, S > 2.39, K < 0.1669$
2	$R > 1.109, D < -0.009, K > 0.1054, RM > 4.654$	$D > -0.115, K < 0.0997, RM > 4.034$
3	$R > 1.029, M > -0.0019, D < -0.056, K > 0.098, RM > 5.21$	$R < 1.024, M > -0.0018, D > -0.065, S < 1.134, K < 0.16$
4	$R > 0.9175, D < -0.009, K > 0.0988, RM > 4.654$	$R < 1.07, D > -0.0426, S < 2.39, K < 0.1669, RM > 4.71$
5	$M > -0.0019, D < -0.0285, K > 0.098895$	$R < 1.029, D > -0.0536, S < 2.39, K < 0.1669$
6	$R > 0.991, M < -0.0019, D < -0.059, K > 0.0988$	$R < 1.078, M < -0.0019, D > -0.068, SK < 2.39, K < 0.166$

The generated patterns illustrate the importance of the proposed methodology because of the clear pattern used to build the cyber-physical maintenance system code. The interpretable methodology illustrates the importance of the Kurtosis and Range factors in the diagnosis of rolling element bearing. Now, the governor parameter needed to build the online remote diagnosis center is present.

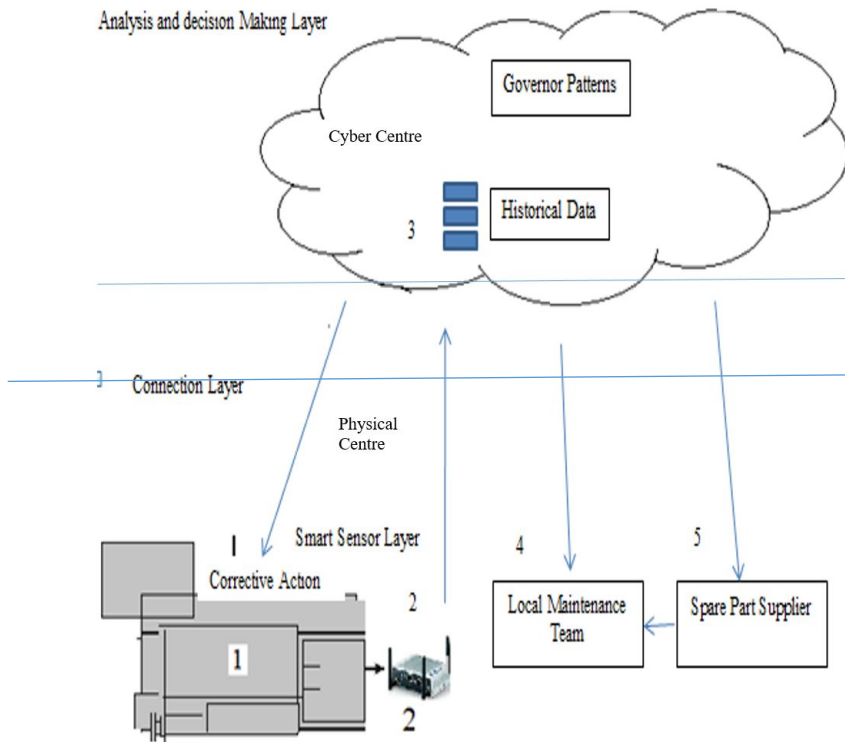


Figure 6. Simulated Cyber-Physical Maintenance System.

As clearly exhibited, CT is the most appropriate means for public communication due to its virtualized resources, parallel processing, security and data service integration with scalable data storage achieving communication between both the cyber world and the physical world. The overall architecture of the proposed system is illustrated in Figure 6. MATLAB was used to build the cyber-physical system code.

The vibration signal is sent from the data set file to the cloud file via the internet in the Dropbox application. It consists of two centers: the physical center, which includes the monitored machines having sensors, actuators, communication Gateway and the local maintenance team.

The cyber center, on the other hand, receives the data coming from the physical system converting it to useful information to monitor and control the physical world without any human interface. The system should also be able to allow the local maintenance team to see a live view of the running machine to interact if required. In the displayed case study, the run to failure data set illustrated in Figure 4 is used. Additionally, MATLAB was used to build the cyber-physical system code.

In our study, part (1) is the aerospace bearing vibration signal data sets that was sent to the Cloud file in Dropbox application via part (2) as infrastructure as a service (IaaS). that is the gateway used to send and receive the signals and corrective action generated from the cyber-physical center via the internet. The function of the gateway device is to bridge the communication gap between the cyber-physical maintenance system layers, the sensors, and the Cloud. It supports communication protocols like Wi-Fi or Bluetooth. For this case study, we used a Wi-Fi gateway. The benefits of using this gateway are: high scalability, lower costs, reduced telecommunication cost and mitigate risks. The built governing code runs every certain time to analyze the received data and generate a suitable decision. For this case study, the software works every 10 minutes, and the period can be adjusted according to the application. In the case of identifying a positive pattern, the system automatically sends an alarm report by secured E-mail to the local maintenance team including the decision and the corrective action. The generated decision is sent back to the machine controller via the gateway carrier. The historical data is then stored in the DropBox file to improve the offline governor patterns generation accuracy. Also, the specification of the spare parts needed for the maintenance process is sent to the spare parts supplier to procure it for the maintenance team.

The generated report, which is sent to the support team, is provided with better understandable patterns as a result of the interpretable methodology. The limitations of the proposed system, on the other hand, are the dependence on the internet service, the real-time online monitoring involving data transfer and the exchange of a large amount of data such as graphs, database information that can cause a time delay and data loss. On a different note, the security of data and anti-hacking technologies is important in the communication field to avoid any unwanted interface. Also, the automatic update of the software by using the stored historical data without an unwanted offline update is an important limitation.

6 Discussion and conclusion

In this paper, a Maintenance-CPS platform built for the diagnosis of machine performance and the exploration of unlabelled data is presented. The advantages of the proposed system are: avoid an expensive onsite visit by expert engineers; reduce training cost and continuous monitoring achieve the near-zero breakdown. The system is based on Cloud technologies to connect the proposed cyber-physical system layers. These layers are a sensor, connection and analysis, and decision making. The analysis and decision making is constructed based on an interpretable methodology dealing with unlabelled data. The proposed methodology consists of three stages; the signal feature extraction, clustering stage using K-mean algorithm and pattern recognition

machine learning technique (LAD). The methodology enables the system to deal with unlabelled data which includes considerable information about monitored machines. Also, the system is tested by using a run to failure data set of aerospace bearing from the NASA prognostic data repository. Samples of patterns generated by using cbmLAD software were shown in Table 4. In summary, the implemented system helps to carry out the maintenance at distance and connects the maintenance chain with near-zero breakdowns and enhances the design of machine element through using Cloud technologies.

In future work, the performance of the cyber-physical maintenance system will be enhanced by improving the proposed technique. A multi-component cyber-physical maintenance system using the same diagnosis center will be the future step. Furthermore, enhancing communication between the system components to increase the reliability of the maintenance task will be considered. This step will be used in real-time monitoring and diagnosis cyber-physical system.

Acknowledgements

The authors would like to thank CAPES for the support provided during this research.

References

- Avram, M. G. (2014). Advantages and challenges of adopting cloud computing from an enterprise perspective. *Procedia Technology*, 12, 529-534. <http://dx.doi.org/10.1016/j.protcy.2013.12.525>.
- Bidgoli, H. (2010). *The handbook of technology management, supply chain management, marketing and advertising, and global management* (Vol. 2). Hoboken: John Wiley & Sons.
- Boros, E., Hammer, P. L., Ibaraki, T., Kogan, A., Mayoraz, E., & Muchnik, I. (2000). An implementation of logical analysis of data. *IEEE Transactions on Knowledge and Data Engineering*, 12(2), 292-306. <http://dx.doi.org/10.1109/69.842268>.
- Campos, J., Sharma, P., Jantunen, E., Baglee, D., & Fumagalli, L. (2016). The challenges of cyber security frameworks to protect data required for the development of advanced maintenance. *Procedia CIRP*, 47, 222-227. <http://dx.doi.org/10.1016/j.procir.2016.03.059>.
- Denkena, B., Bluemel, P., Kroening, S., & Roebbing, J. (2012). Condition based maintenance planning of highly productive machine tools. *Production Engineering*, 6(3), 277-285. <http://dx.doi.org/10.1007/s11740-011-0351-9>.
- Dhillon, B. S. (2006). *Maintainability, maintenance, and reliability for engineers*. Boca Raton: CRC Press. <http://dx.doi.org/10.1201/9781420006780>.
- Efthymiou, K., Papakostas, N., Mourtzis, D., & Chryssolouris, G. (2012). On a predictive maintenance platform for production systems. *Procedia CIRP*, 3, 221-226. <http://dx.doi.org/10.1016/j.procir.2012.07.039>.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1), 10-18. <http://dx.doi.org/10.1145/1656274.1656278>.
- Hammer, P. L. (1986). Partially defined Boolean functions and cause-effect relationships. In *Proceedings of the International Conference on Multi-Attribute Decision Making via or-based Expert Systems*. Passau: Universität Passau.

- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of "big data" on cloud computing: review and open research issues. *Information Systems*, 47, 98-115. <http://dx.doi.org/10.1016/j.is.2014.07.006>.
- Holgado, M., Macchi, M., & Fumagalli, L. (2016). Value-in-use of e-maintenance in service provision: survey analysis and future research agenda. *IFAC-Papers On Line*, 49(28), 138-143. <http://dx.doi.org/10.1016/j.ifacol.2016.11.024>.
- Jantunen, E., Gorostegui, U., Zurutuza, U., Albano, M., Ferreira, L. L., Hegedús, C., & Campos, J. (2018). Remote maintenance support with the aid of cyber-physical systems and cloud technology. *Proceedings of the Institution of Mechanical Engineers. Part I, Journal of Systems and Control Engineering*, 232(6), 784-794. <http://dx.doi.org/10.1177/0959651818772939>.
- Jantunen, E., Zurutuza, U., Ferreira, L. L., & Varga, P. (2016). Optimising maintenance: what are the expectations for cyber physical systems. In *Proceedings of the 3rd International Workshop on Emerging Ideas and Trends in Engineering of Cyber-Physical Systems (EITEC)* (pp. 53-58). Piscataway: IEEE.
- Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient K-means clustering algorithm: analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 881-892. <http://dx.doi.org/10.1109/TPAMI.2002.1017616>.
- Kirubashankar, R., & Krishnamurthy, K. (2013). A real-time web-enabled platform for information monitoring and fault diagnosis in a distributed control system. *Instrumentation Science & Technology*, 41(3), 236-250. <http://dx.doi.org/10.1080/10739149.2012.735309>.
- Lakis, A. A. (2007). Rotating machinery fault diagnosis using time-frequency methods. In *Proceedings of the 7th WSEAS International Conference on Electric Power Systems, High Voltages, Electric Machines*. Athens: WSEAS Press.
- Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters*, 3, 18-23. <http://dx.doi.org/10.1016/j.mfglet.2014.12.001>.
- Linderoth, J. T., & Lodi, A. (2010). MILP software. In J Cochran (Ed.), *Wiley encyclopedia of operations research and management science*. New York: Wiley.
- Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G., Sihn, W., & Ueda, K. (2016). Cyber-physical systems in manufacturing. *CIRP Annals*, 65(2), 621-641. <http://dx.doi.org/10.1016/j.cirp.2016.06.005>.
- Moreira, L. M. (2000). *The use of Boolean concepts in general classification contexts* (Doctoral dissertation). Ecole Polytechnique Federale de Lausanne, Lausanne.
- Mori, M., Fujishima, M., Komatsu, M., Zhao, B., & Liu, Y. (2008). Development of remote monitoring and maintenance system for machine tools. *CIRP Annals-Manufacturing Technology*, 57(1), 433-436. <http://dx.doi.org/10.1016/j.cirp.2008.03.108>.
- Mortada, M. A., Yacout, S., & Lakis, A. (2014). Fault diagnosis in power transformers using multi-class logical analysis of data. *Journal of Intelligent Manufacturing*, 25(6), 1429-1439. <http://dx.doi.org/10.1007/s10845-013-0750-1>.
- Mortada, M.-A., Yacout, S., & Lakis, A. (2011). Diagnosis of rotor bearings using logical analysis of data. *Journal of Quality in Maintenance Engineering*, 17(4), 371-397. <http://dx.doi.org/10.1108/13552511111180186>.
- Mourtzis, D., Vlachou, A., & Zogopoulos, V. (2017). Cloud-based augmented reality remote maintenance through shop-floor monitoring: a product-service system approach. *Journal of Manufacturing Science and Engineering*, 139(6), 061011. <http://dx.doi.org/10.1115/1.4035721>.
- Mourtzis, D., Vlachou, E., Milas, N., & Xanthopoulos, N. (2016). A cloud-based approach for maintenance of machine tools and equipment based on shop-floor monitoring. *Procedia CIRP*, 41, 655-660. <http://dx.doi.org/10.1016/j.procir.2015.12.069>.

- Qiu, H., Lee, J., Lin, J., & Yu, G. (2006). Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. *Journal of Sound and Vibration*, 289(4), 1066-1090. <http://dx.doi.org/10.1016/j.jsv.2005.03.007>.
- Ryoo, H. S., & Jang, I. Y. (2009). Milp approach to pattern generation in logical analysis of data. *Discrete Applied Mathematics*, 157(4), 749-761. <http://dx.doi.org/10.1016/j.dam.2008.07.005>.
- Sanchez, C. R., Fernandez, F. J. G., Vena, L. C. S., Carpio, J., & Castro, M. (2011). Industrial telemaintenance: remote management experience from subway to industrial electronics. *IEEE Transactions on Industrial Electronics*, 58(3), 1044-1051. <http://dx.doi.org/10.1109/TIE.2010.2049705>.
- Sanz-Bobi, M. A., Roque, A. M. S., Marcos, A., & Bada, M. (2012). Intelligent system for a remote diagnosis of a photovoltaic solar power plant. *Journal of Physics: Conference Series*, 364(1), 012119. <http://dx.doi.org/10.1088/1742-6596/364/1/012119>.
- Sekar, R., Hsieh, S. J., & Wu, Z. (2011). Remote diagnosis design for a PLC-based automated system: 1-implementation of three levels of architectures. *International Journal of Advanced Manufacturing Technology*, 57(5), 683-700. <http://dx.doi.org/10.1007/s00170-011-3314-6>.
- Shaban, Y., Meshreki, M., Yacout, S., Balazinski, M., & Attia, H. (2017). Process control based on pattern recognition for routing carbon fiber reinforced polymer. *Journal of Intelligent Manufacturing*, 28(1), 165-179. <http://dx.doi.org/10.1007/s10845-014-0968-6>.
- Venkataraman, V. (2010). *Maintenance engineering and management*. New Delhi: PHI Learning Private Limited.
- Villars, R. L., & Olofson, C. W., & Eastwood, M. (2011). *Big data: what it is and why you should care* (White Paper). USA: IDC.
- Wang, J. F., Tse, P. W., He, L. S., & Yeung, R. W. (2004). Remote sensing, diagnosis and collaborative maintenance with web-enabled virtual instruments and mini-servers. *International Journal of Advanced Manufacturing Technology*, 24(9-10), 764-772. <http://dx.doi.org/10.1007/s00170-003-1753-4>.
- Wang, J., Zhang, L., Duan, L., & Gao, R. X. (2017). A new paradigm of cloud-based predictive maintenance for intelligent manufacturing. *Journal of Intelligent Manufacturing*, 28(5), 1125-1137. <http://dx.doi.org/10.1007/s10845-015-1066-0>.
- Wang, L. (2013). Machine availability monitoring and machining process planning towards cloud manufacturing. *CIRP Journal of Manufacturing Science and Technology*, 6(4), 263-273. <http://dx.doi.org/10.1016/j.cirpj.2013.07.001>.
- Yacout, S. (2010). Fault detection and diagnosis for condition based maintenance using the logical analysis of data. In *Proceedings of the 40th International Conference on Computers and Industrial Engineering*. Piscataway: IEEE. <http://dx.doi.org/10.1109/ICCIE.2010.5668357>.
- Yacout, S., Salamanca, D., & Mortada, M.-A. (2012, January 26). Tool and method for fault detection of devices by condition based maintenance. *Patent Cooperation Treaty PCT/CA2011/000876 WO/2012/009804*.
- Zhang, K. (2008). *Visual cluster analysis in data mining* (Thesis PhD). Macquarie University, Australia.
- Zhao, F., Chen, J., Dong, G., & Guo, L. (2010). SOA-based remote condition monitoring and fault diagnosis system. *International Journal of Advanced Manufacturing Technology*, 46(9), 1191-1200. <http://dx.doi.org/10.1007/s00170-009-2178-5>.